WE CLAIM:

1.   A method for identifying a drug discovery target which comprises:

(a)   providing a means for accessing genomics information in a database wherein said means permits computational analysis of biological relationships among the stored concepts;

(b)   generating one or more subsets of genomics information from the database wherein at least one of the one or more subsets is a disease-related pathway; and

(c)   identifying the biological interactions and actor concepts in the disease-related pathway whereby each of the actor concepts involved in each such reaction is a drug discovery target.

2.   The method of claim 1 wherein the genomics information comprises information relating to genes, their DNA sequences, mRNA, the proteins that result when the genes are expressed, and the biological effects of the expressed proteins.

3.   The method of claim 2 wherein the data comprise data extracted from multiple public sources.

4.   The method of claim 2 wherein the data comprises proprietary data.

5.   The method of claim 2 wherein the data comprises data extracted from a combination of proprietary and public data sources.

6.   The method of claim 2 wherein the data comprises data extracted from a combination of proprietary and public data sources.

7.   The method of claim 2 wherein the means for storing the genomics information includes an ontology in which:

(a)   each gene, gene product, and biological effect is given an identifier which is related to synonyms for the identifier;

(b)   each gene, gene product, and biological effect is categorized by class; and

1-WA/2030998.1

(c)   the relationship of each gene, gene product and disease state is defined by slots and facets.

8.   The method of claim 2 wherein the candidate drug discovery targets in the disease related pathway are prioritized based on factors that include function and complexity.

9.   The method of claim 8 wherein the candidate drug discovery targets are further prioritized based on markers for side effects and patient responsiveness.

10.   The method of claim 2, further comprising combining the results of querying the database with the results of additional data obtained from one or more additional methods for identifying candidate drug discovery targets.

11.   The method of claim 10 wherein the additional data is obtained from one of protein-protein interaction studies and protein profiling from mass spectrometry.

12.   The method of claim 11 wherein the additional data is obtained from differential gene expression studies.

13.   The method of claim 1 wherein the genomics information comprises information relating to genotype and the disease-related pathway comprises a gene or gene product associated with a particular genotype.

14.   The method of claim 1 wherein the genomics information comprises the name of each gene, gene product, and their biological effects, and the means for storing and accessing the genomics information identifies relationships that are at least one step removed.

15.   The method of claim 1, wherein the identifying the biological interactions step includes the step of comparing genomics data from the database to user-defined data using a statistical model.

16.   The method of claim 15, wherein the comparing step includes identifying an overlap between user-defined data and data from the database and the statistical model is a statistical significance model measuring the likelihood that the overlap is a random event.

17.   The method of claim 16, wherein the user-defined data is one of gene expression data and a manually entered gene list.

18.   The method of claim 1, wherein the identifying the biological interactions step includes classifying one or more relevant findings using an ontology.

19.   The method of claim 18, wherein the classifying one or more findings using an ontology includes determining a likelihood that the one or more findings residing in a particular biological classification in the ontology is statistically significant.

20.   The method of claim 1, wherein the generating one or more subsets of genomics information step includes the step of generating profiles according to one or more criteria.

21.   The method of claim 20, wherein the profiles are pre-generated from the database.

22.   The method of claim 20, wherein the profiles are generated by one of a data-driven and model-driven approach.

23.   The method of claim 20, wherein the profiles are generated based upon information contained in the database and user supplied genomics information.

24.   The method of claim 20, wherein the generating of profiles includes generating a locus for the one or more subsets based upon one of the data contained in the database, the user-supplied genomics data, or a combination of the data contained in the database and the user-supplied genomics data.

26.   The method of claim 20, further comprising the step of providing a user-supplied set of gene expression data for identifying a particular subset of genomics information, wherein the gene expression data are selected based on one or more of expression levels derived from a microarray experiments   a prior analysis algorithm, and a user's preferred gene set.

27.   The method of claim 20, wherein the profiles are gene-centric being derived about a central gene for all genes in the database.

28.   The method of claim 20, wherein the generating the profiles step includes deriving profiles about one or more sets of related user-selected genes.

29.   The method of claim 28, wherein the profiles are generated so as to be non-overlapping by ensuring that user-selected genes do not appear in more than a predetermined maximum threshold number of generated profiles.

30.   The method of claim 28, wherein the profiles are generated so as to be based on connections between a first known drug target gene and a second gene of interest.

31.   The method of claim 20, wherein the profiles are at least one of process, function and disease centric being derived about a central biological process for all processes in the database.

32.   The method of claim 20, wherein the profiles are at least one of tissue, organ or structure centric being derived about a central physical object for all objects in the database.

33.   A method for evaluating user-supplied genomics data using a structured database that permits the computation of complex relationships among genes and/or gene products contained in the database, comprising:
(a)   defining a profile model based on one or more profile definition criterion;
(b)   building a collection of profiles according to the profile model;

(c)   identifying one or more profiles that overlap at least a portion of the user-supplied genomics data and determining, for each such overlapped profile, whether the overlap is statistically significant; and

(d)   analyzing one or more statistically significant profiles together with the user-supplied genomics data including inspecting database-asserted biological interactions embodied in the one or more statistically significant profiles.

34.   The method of claim 33, further including the step of pre-generating a profile library containing a profile for each one of a genomic information type in the database according to the profile model.

35.   The method of claim 33, wherein the defining the profile model includes the step of selecting a profile seed according to one or more attributes of the user-supplied genomics data.

36.   The method of claim 34, wherein the profile seed is a user-supplied differentially expressed gene.

37.   The method of claim 34, wherein the profiles are pre-generated from a graph structure and user-supplied genomics data.

38.   The method of claim 33, further comprising the step of generating profiles by querying the database for information matching the one or more profile definition criterion.

39.   The method of claim 33, wherein the determining statistical significance step includes the step of computing a probability of overlap as a function of information contained in the database.

40.   The method of claim 33, wherein the genomic information type is one of a gene, gene product and biological process.

41.   The method of claim 33, wherein the user-supplied genomics data is differential gene expression data and the analyzing step further includes one of the steps of:

1-WA/2030998.1

(1)   identifying a new use for a known therapy wherein the gene expression data relates to a pathway affected by the known therapy;

(2)   prioritizing candidate development compounds for further development wherein the gene expression data relates to the target of one or more candidate development compounds and the analyzing step includes giving higher priority to development compounds on the basis of whether or not they are likely to result in an undesirable effect based on their involvement in other biological pathways as embodied in the profile; and

(3)   identifying disease-related pathways wherein the disease is a side effect of drug therapy, wherein the gene expression data relates to the target affected by the drug therapy and the alternative pathways that are also affected by the drug or the drug discovery target and that result in an undesirable phenotype are embodied in the profile.

42.   The method of claim 33, wherein the genomics data is differential gene expression data relating to particular disease, and wherein the analyzing step further includes the step of validating whether the gene expression data are genotypic markers for the disease state according to whether a database-asserted biological association related to the disease state, which is shared among a plurality of overlapped profiles, is statistically significant.

43.   The method of claim 33, wherein the profile generation criterion include one or more of a biological process, number of genes, organismal, gene connectivity, edge connectivity, findings source type, experiment context, and tissue consistency criterion.

44.   The method of claim 33, wherein the profiles are generated from a seed node and the inspecting database-asserted biological interactions step focuses on the biological interactions emanating from the seed node.

45.   The method of claim 44, wherein the seed is one of a gene, gene product and biological process genomic data type.

46.   The method of claim 33, further comprising computing a statistical significance for a biological association in the one or more statistically significant profiles.

47. The method of claim 33, wherein the generating a profile library step includes, for each profile generated, the step of selecting a node for a profile based on the number of similar findings in the database that link the node to a neighboring node.

48. The method of claim 33, further comprising the step of displaying information related to the one or more statistically significant profiles and genomics data using a GUI.

49. The method of claim 33, further including the step of annotating the profiles with biological associations asserted by the database including one or more of a cellular process, molecular process, organismal process and disease process.

50. The method of claim 49, further including the step of displaying biological association using one of a GUI and a report.

51. The method of claim 49, wherein the annotation of profiles includes using classification information found in an ontology.

52. The method of claim 33, wherein the determining of statistical significance test includes testing a null hypothesis over a discrete probability distribution, the distribution being a function of the database size, profile sizes, the user-supplied genomics data size and expression values.

53. The method of claim 33, wherein the generating step includes generating a plurality of profile libraries, each of which corresponding to a different one of a plurality of profile generation criterions.

54. A method for identifying a new use for a known therapy comprising the steps of providing a means for accessing genomics information in a database wherein said means permits computational analysis of biological relationships among the stored concepts; generating subsets of genomics information from the database wherein the subsets include disease-related pathways comprising a known therapy target; selecting at least one of such disease-related

pathways wherein the known therapy target is also comprised within a second disease-related pathway; and identifying a treatment of the second disease as a new use for the known therapy.

55. A method for prioritizing candidate development compounds for further development that comprises the steps of providing a means for accessing genomics information in a database wherein said means permits computational analysis of biological relationships among the stored concepts; generating one or more subsets of genomics information from the database; and querying the one or more subsets to identify all pathways associated with the target of each candidate development compound and assigning higher priority to development compounds on the basis of whether or not they are likely to result in an undesirable effect based on their involvement in other biological pathways.

56. A method for identifying disease-related pathways wherein the disease is a side effect of drug therapy that comprises the steps of providing a means for accessing genomics information in a database wherein said means permits computational analysis of biological relationships among the stored concepts; identifying a disease-related pathway affected by a drug or drug discovery target; generating one or more subsets of genomics information from the database relating to the disease-related pathway; and querying the one or more subsets to identify alternative pathways that are also affected by the drug or the drug discovery target and that result in the undesirable phenotype.

57. A method for identifying or validating a genotypic marker for a disease state that comprises providing a means for accessing genomics information in a database wherein said means permits computational analysis of biological relationships among the stored concepts; generating one or more subsets of genomics information from the database; and querying the one or more subsets to identify a genotypic marker that is associated with the disease state.

58. A method for identifying a drug discovery target which comprises querying one or more subsets of genomics information generated from a database to identify a disease-related pathway, the biological interactions and actor concepts in the disease-related pathway, whereby at least one of the actor concepts involved in each such reaction is a drug discovery target and wherein

the one or more subsets reside on a computer system comprising a means for accessing genomics information in the database and a means for computational analysis of biological relationships among the concepts contained within the subsets.

59.   A method of conducting business that comprises receiving compensation from a customer in return for identifying to the customer a drug discovery target discovered by querying a computer system to identify a disease-related pathway and identifying the biological interactions and actor concepts in the disease-related pathway whereby each of the actor concepts involved in each such reaction is a drug discovery target and wherein the computer system comprises a means for accessing genomics information in a database, generating subsets of genomics information from the database, and a means for computational analysis of biological relationships among the concepts represented in the subsets.

60.   A drug discovery target identified by the process of claim 1.

61.   A method of drug discovery that comprises:
(a) identifying a drug discovery target discovered by querying a computer system to identify a disease-related pathway and identifying the biological interactions and actor concepts in the disease-related pathway whereby one or more of the actor concepts involved in each such reaction is a drug discovery target and wherein the computer system comprises a means for computing subsets of genomics information specific to one or more target diseases using a database of genomics information, and a means for computational analysis of biological relationships among the concepts included within the subsets; and
(b) screening compounds against a drug discovery target to identify drug candidates.

62.   A drug candidate identified by the process of claim 61.